

Independent uncertainty estimates for coefficient based sea surface temperature retrieval from the Along-Track Scanning Radiometer instruments

Article

Accepted Version

Bulgin, C. E., Embury, O., Corlett, G. and Merchant, C. J. (2016) Independent uncertainty estimates for coefficient based sea surface temperature retrieval from the Along-Track Scanning Radiometer instruments. *Remote Sensing of Environment*, 178. pp. 213-222. ISSN 0034-4257 doi: <https://doi.org/10.1016/j.rse.2016.02.022> Available at <https://centaur.reading.ac.uk/57628/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1016/j.rse.2016.02.022>

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Independent uncertainty estimates for coefficient based sea surface temperature retrieval from the Along-Track Scanning Radiometer instruments

C. E. Bulgin^a, O. Embury^a, G. Corlett^b, C. J. Merchant^a

^a*Department of Meteorology, University of Reading, Reading, UK*

^b*University of Leicester, Department of Physics & Astronomy, University of Leicester, UK*

Abstract

We establish a methodology for calculating uncertainties in sea surface temperature estimates from coefficient based satellite retrievals. The uncertainty estimates are derived independently of in-situ data. This enables validation of both the retrieved SSTs and their uncertainty estimate using in-situ data records. The total uncertainty budget is comprised of a number of components, arising from uncorrelated (eg. noise), locally systematic (eg. atmospheric), large scale systematic and sampling effects (for gridded products). The importance of distinguishing these components arises in propagating uncertainty across spatio-temporal scales. We apply the method to SST data retrieved from the Advanced Along Track Scanning Radiometer (AATSR) and validate the results for two different SST retrieval algorithms, both at a per pixel level and for gridded data. We find good agreement between our estimated uncertainties and validation data. This approach to calculating uncertainties in SST retrievals has a wider application to data from other instruments and retrieval of other geophysical variables.

Keywords:

1 **1. Introduction**

2 Uncertainty is inherent in all geophysical measurements and must be ap-
3 propriately characterised for their scientific application. Data providers have
4 a responsibility to communicate the levels of uncertainties associated with
5 their products and inform data users of the correct methodology for using
6 uncertainty information provided. Within the Sea Surface Temperature Cli-
7 mate Change Initiative (SST CCI) project (Hollmann et al., 2013; Merchant
8 et al., 2014) we aim to provide an uncertainty budget for every SST value
9 provided in products (skin temperature, SST at 0.2 m depth and spatially
10 averaged SST). We aim to derive uncertainty estimates independently of SST
11 validation datasets, allowing validation of both the SST values and their as-
12 sociated uncertainty.

13 The terms ‘error’ and ‘uncertainty’ are sometimes used interchangeably,
14 but have distinct standard definitions that will be adhered to throughout this
15 paper. Error is the difference between a measured value and the true value of
16 the measurand (JCGM, 2008; Kennedy, 2014). In practice we know neither
17 the true value nor therefore the error for a particular measurement. However
18 the distribution of the errors can often be estimated and this distribution
19 characterises the uncertainty in the measured value. Formally, uncertainty
20 is a parameter characterising the dispersion of values that could reasonably
21 be attributed to the measured value (JCGM, 2008). To quantify uncertainty
22 in this paper we quote one standard deviation of the error distribution.

23 It is common to provide generic uncertainty estimates for remotely sensed
 24 SST derived via comparison with in-situ datasets during validation activities.
 25 The standards of the Group for High Resolution Sea Surface Temperature
 26 (GHRSSST) specify the provision in all datasets of single sensor error statis-
 27 tics (SSES). For pragmatic reasons, SSES are defined to comprise the mean
 28 difference and standard deviation of remotely sensed SST matched to a ‘refer-
 29 ence’ dataset (GHRSSST Science Team, 2010). Drifting buoy SSTs are often
 30 used as the ‘reference’. Mean and standard deviation validation statistics
 31 are often provided as globally invariant dataset specific values (May et al.,
 32 1997; Reynolds et al., 2002; Casey and Cornillon, 1998). An additional field
 33 indicating the retrieval quality level can be specified at pixel resolution pro-
 34 viding information on the likelihood of cloud contamination, noise amplifi-
 35 cation at extreme satellite zenith angles or input data quality (Donlon et al.,
 36 2007; Kilpatrick et al., 2001). An extension of this approach is the MOD-
 37 erate Resolution Infrared Spectrometer (MODIS) algorithm, which provides
 38 validation-based uncertainty information stratified by season, latitude, sur-
 39 face temperature, satellite zenith angle, a selected brightness temperature
 40 difference, SST quality level and day/night (Castro et al., 2010).

41 Sources of uncertainty in remotely sensed SST are intrinsic to the retrieval
 42 process and the data utilised. Uncertainties vary from pixel to pixel due to
 43 local changes in instrument noise, satellite viewing geometry and atmospheric
 44 conditions. We present here a method of estimating SST retrieval uncertainty
 45 that accounts for these factors at the pixel level. There are a number of
 46 sources of uncertainty in SST measurement and the need to differentiate the
 47 effects of random, and systematic errors has been previously noted (Reynolds

et al., 2002; Casey and Cornillon, 1998; Merchant et al., 2012; Kennedy,
 2014). Gridding of products introduces sampling uncertainties and a number
 of studies have considered these when constructing global or regional SST
 datasets from in-situ observations (She et al., 2007; Folland et al., 2001;
 Rayner et al., 2006; Morrissey and Greene, 2009; Jones et al., 1997; Brohan
 et al., 2006).

In this paper, we consider uncorrelated and locally systematic effects con-
 tributing to SST uncertainty. The random or uncorrelated effects arise from
 noise in the satellite brightness temperature, which propagates into the re-
 trieved SST. Locally systematic effects cause errors that are correlated on
 synoptic scales of atmospheric variability and are related to the retrieval
 method itself interacting with changes in atmospheric properties (Minnett,
 1991; Barton, 1998; Le Borgne et al., 2011; Minnett and Corlett, 2012;
 Embury and Merchant, 2012; Merchant et al., 2012). We also discuss un-
 certainties from large scale systematic effects (spatially coherent on larger
 scales than synoptic features). In a companion paper (Bulgin et al., 2016)
 we derive a method for calculating sampling uncertainty in gridded products
 due to incomplete sampling of observations in each grid cell, primarily as a
 result of cloud cover. In this paper, we use results from Bulgin et al. (2016),
 and, for completeness, show how sampling uncertainty combines with other
 components of uncertainty in gridded products.

The remainder of the paper is structured as follows. Section 2 describes
 the theory behind the calculation of uncertainties, their propagation and how
 this is applied to different levels of SST data (orbit data and gridded prod-
 ucts). Section 3 describes how an initial uncertainty budget is constructed

73 from errors originating from random, locally correlated and sampling effects.
74 In Section 4 we present a validation of our uncertainty budget and in Section
75 5 provide a discussion of the results. We conclude the paper in Section 6.

76 2. Uncertainty Calculation and Propagation

77 We construct an uncertainty budget for SST measurements in CCI prod-
78 ucts comprised of uncertainty components arising from random, locally sys-
79 tematic, large-scale systematic and sampling effects. The full equation for
80 the propagation of uncertainty in a variable y , ($u(y)$), given that y is related
81 to input quantities x_i via $y = f(x_1, \dots, x_n)$, is defined as equation (1) in the
82 Guide to the Expression of Uncertainty in Measurement (GUM) (JCGM,
83 2008).

$$u^2 = \sum_i^n \left(\frac{\partial f}{\partial x_i} \right)^2 u_i^2(x_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(\frac{\partial f}{\partial x_i} \right) \left(\frac{\partial f}{\partial x_j} \right) u(x_i, x_j) \quad (1)$$

84 Uncertainty is expressed with respect to (y) in the GUM, and we repro-
85 duce this notation throughout the paper. However, in Earth Observation,
86 we conventionally relate a retrieval estimate \hat{x} to observations y ie. $\hat{x} = f(y)$
87 which is the reverse convention. The first term in equation (1) describes the
88 propagation of uncertainties from uncorrelated errors. These can be added
89 in quadrature with the differential term ($\partial f / \partial x_i$) defining the sensitivity of
90 the total uncertainty to each uncertainty component. The second term de-
91 scribes the propagation of uncertainty terms arising from correlated errors.
92 This term sums the uncertainty components from correlated errors for each
93 pair of input variables (x_i and x_j) found as the product of the sensitivity for

94 both x_i and x_j and the covariance between them, $u(x_i, x_j)$. The factor of ‘2’
 95 is included, as for each pair, each is equally correlated with the other.

96 Equation (1) can also be written in the form of equation (2) where the
 97 uncertainty is expressed as the sum over all pairs of input variables and the
 98 covariance term is expressed as the product of the standard uncertainty in
 99 x_i , written u_i , in x_j , written u_j , and of the correlation of errors in x_i and x_j ,
 100 written r_{ij} .

$$u^2 = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} u_i u_j r_{ij} \quad (2)$$

101 Equation (2) applies fairly generically to any transformation $y = f(x_i, \dots, x_n)$
 102 for which the sensitivity parameters $(\partial f / \partial x_i)$ are adequately constant over
 103 the range $x_i - u_i$ to $x_i + u_i$; it is a first order approximation. Because we
 104 will use the results later, we illustrate the use of equation (2) for calculat-
 105 ing the uncertainty in the mean SST from a number of observations. If
 106 $f = \sum_{i=1}^n x_i / n$, where each x_i is a contributing SST value, then the sensitiv-
 107 ity parameter is $\partial f / \partial x_i = 1/n$ giving:

$$u^2 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n u_i u_j r_{ij} \quad (3)$$

108 We can consider three limiting cases. First assume errors are uncorrelated
 109 between pixels. We can then put $r_{ij} = \delta_{ij}$, where $\delta_{ij} = 1$ for $i = j$, and $\delta_{ij} = 0$
 110 for $i \neq j$. In this case, the uncertainty in the mean is scaled by the familiar
 111 ‘ $\frac{1}{\sqrt{n}}$ ’ reduction in uncertainty, because

$$u^2 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n u_i u_j \delta_{ij} \quad (4)$$

$$= \frac{1}{n^2} \sum_i^n u_i^2 \quad (5)$$

112 Second, consider the case $r_{ij} = 1$, which means errors fully correlate
 113 between contributing SSTs. Equation (3) becomes

$$u^2 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n u_i u_j \quad (6)$$

$$= \frac{1}{n^2} \left(\sum_{i=1}^n u_i \right)^2 \quad (7)$$

114 implying $u = \frac{1}{n} \sum_{i=1}^n u_i$ ie. the uncertainty is the average uncertainty of
 115 the contributing SSTs.

116 Third, consider the case $r_{ij} = \delta_{ij} + (1 - \delta_{ij})r$ - all SSTs have the same
 117 error correlation with other SSTs. Substituting into equation (3) gives

$$u^2 = \frac{1}{n^2} \sum_i^n \sum_j^n u_i u_j [\delta_{ij} + (1 - \delta_{ij})r] \quad (8)$$

$$= \frac{1}{n^2} \sum_i^n \sum_j^n u_i u_j [r + (1 - r)\delta_{ij}] \quad (9)$$

$$= \frac{r}{n^2} \left(\sum_{i=1}^n u_i \right)^2 + \frac{(1-r)}{n^2} \left(\sum_{i=1}^n u_i^2 \right) \quad (10)$$

118 This form yields the previous results as special cases ($r = 0$ and $r =$
 119 1). Constant r_{ij} for $i \neq j$ is in practice unlikely to be exact for a real
 120 situation, but may be a useful approximation in some cases, avoiding the
 121 need to estimate r_{ij} for every contributing pair.

122 3. Uncertainty Budget Components

123 3.1. Uncorrelated Effects

124 Random errors in SST estimation from satellite data arise from noise
125 in the satellite observations. The signal recorded by a typical radiometer
126 is a voltage measured across a detector, digitised and recorded as counts.
127 In the operational calibration, a linear radiance is calculated in the form
128 $radiance = gain \times counts + offset$ where the gain and count parameters are
129 calculated during instrument calibration (Smith et al., 2012). A non-linearity
130 adjustment is then applied to the longwave channels (Smith et al., 2012) for
131 which the associated uncertainty has not been calculated. In this analysis
132 we simply take the detector noise in the measured counts and propagate this
133 into our geophysical retrieval. In a coefficient based retrieval, SST is calcu-
134 lated from a pre-defined linear or nearly linear (Anding and Kauth (1970);
135 Deschamps and Phulpin (1980); Kilpatrick et al. (2001); May et al. (1997);
136 McMillan and Crosby (1984), and further references within Merchant (2013))
137 combination of the observed brightness temperatures. Brightness tempera-
138 ture uncertainty is characterised using channel-specific noise equivalent dif-
139 ferential temperature (NEdT). This uncertainty is then propagated into the
140 SST retrieval uncertainty.

141 We illustrate the propagation of errors from random effects using data
142 from the polar orbiting Advanced Along Track Scanning Radiometer (AATSR)
143 aboard the Envisat satellite. Envisat was in a sun synchronous orbit with
144 an equator overpass time of 10.00 am. AATSR made observations in seven
145 spectral bands covering the visible and infrared spectrum at two viewing ge-
146 ometries: nadir ($0 - 22^\circ$) and forward ($52 - 55^\circ$). SST can be derived using

147 the nadir infrared channels only, or using both the nadir and forward views.
 148 We consider here the propagation of uncertainties through two different re-
 149 trievals: ‘N2’ using the 11 and 12 μm channels in the nadir view only and
 150 ‘D2’ using the 11 and 12 μm channels in both views. The formula used here
 151 for estimating coefficient based SSTs from satellite data is:

$$\hat{x}_{SST} = a_0 + \sum_k a_k y_k \quad (11)$$

152 Where y_k refers to each channel used in the retrieval, a_0 is an offset and a_k
 153 are channel specific coefficients. Note that here $\hat{x} = f(y)$, in contrast to usage
 154 in Section 2 (as previously noted). These coefficients vary with the context in
 155 which the observation is made, according to the viewing geometry and total
 156 column water vapour (TCWV), but are predefined. The error (difference
 157 between the measured value and true value) for a given SST can be defined
 158 as:

$$e_{SST} = \sum_k a_k e_{y_k} \quad (12)$$

159 This is a linear combination of the errors in the brightness temperatures
 160 in each channel (denoted by ‘k’) multiplied by the coefficient used in the
 161 retrieval. In practice, we do not know the true SST value nor therefore the
 162 error on each individual measurement, but we can simulate a ‘typical’ error
 163 field from our knowledge of the NEdT in each channel. We illustrate this in
 164 panels 1 and 2 of Figure 1 which show simulated error fields for the nadir
 165 view of the 11 and 12 μm channel at pixel resolution (1 km at nadir for
 166 AATSR). These are constructed using a Gaussian random number generator
 167 selecting values from a distribution with 0.0 $^{\circ}\text{C}$ mean and 0.05 $^{\circ}\text{C}$ standard

Table 1: Coefficients for each channel used to calculate SST in the ‘N2’ and ‘D2’ retrievals in Figure 1.

Retrieval	Channel	Sec(Sat Zenith Angle)	a_1 coefficient
N2	11 μm	1.0	2.04314
N2	12 μm	1.0	-1.02542
D2	11 μm	1.0	4.65371
D2	11 μm	1.76	-1.65009
D2	12 μm	1.0	-3.27043
D2	12 μm	1.76	1.27186

168 deviation representing NEdT estimates for the two channels (Embury and
169 Merchant, 2012). Errors vary in magnitude from pixel to pixel and can be
170 either positive or negative in sign.

171 Panels 3 and 4 of Figure 1 show the propagation of these simulated error
172 fields in a N2 and D2 retrieval. For the purpose of this illustration we assume
173 a fixed view angle and TCWV (23 kg m^{-2}) across the image giving coefficients
174 (a_k) dependent only on channel, as shown in Table 1. Under normal retrieval
175 conditions these would vary slightly on a per-pixel basis. The coefficients
176 are specified to five decimal places (Merchant and LeBorgne, 2004). Further
177 discussion of error inherent in the retrieval process is provided in Section
178 3.2. As indicated in equation (12) the uncorrelated errors in a given retrieval
179 are the sum of the errors in each channel, and therefore the total errors are
180 smaller in the N2 retrieval than the D2 retrieval (which uses four channels
181 with generally heavier weights).

182 Many users require gridded Level 3 products generated from full reso-

183 lution data. When generating gridded products, the average SST can be
 184 calculated using the arithmetic mean:

$$\hat{x}_{GriddedSST} = \frac{1}{n} \sum_{i=1}^n \hat{x}_{SST(i)} \quad (13)$$

185 Where n is the number of observations (i) in the grid cell. The alternative
 186 would be to calculate a weighted mean based on the per-pixel uncertainties,
 187 but we choose the arithmetic mean as it gives equal weight to all measure-
 188 ments across the grid cell and therefore represents a mean across the geo-
 189 physical variability within the grid cell. Panels 5 and 6 show the arithmetic
 190 mean of the errors over a 5 x 5 pixel grid cell, approximately representing the
 191 creation of 0.05° Level 3 products. The range in the mean error is naturally
 192 smaller in the gridded product, but remains larger for the D2 retrieval than
 193 the N2 retrieval.

194 In practice, when retrieving SST from satellite observations we don't ex-
 195 plicitly know the error in either the brightness temperatures or SST. We need,
 196 however, to estimate the uncertainty in the SST retrieval. Given estimates
 197 of NEdT, this is an example of standard uncertainty propagation. 'Standard
 198 uncertainty' is the standard deviation of errors in each channel brightness
 199 temperature, estimated to be of the order of 0.05 K for both the 11 and 12
 200 μm channels of AATSR (Embury and Merchant, 2012). The propagation of
 201 uncorrelated uncertainty components is shown in equation (5) where uncer-
 202 tainties are added in quadrature. Applying this to equation (11), in the first
 203 instance to give the per pixel uncertainty, and differentiating with respect to
 204 each channel (y_k) used in the retrieval gives:

$$u_i = \sqrt{\sum_k a_k^2 u_{y_k}^2} \quad (14)$$

205 For a gridded product using the arithmetic mean, the uncertainty in the
 206 mean of the contributing pixels is

$$u_{GriddedSST} = \frac{1}{\sqrt{n}} \sqrt{\frac{\sum u_i^2}{n}} \quad (15)$$

207 For fixed coefficients and a constant error in the brightness temperatures
 208 (0.05 K) as in Figure 1, there is an invariant uncertainty value for each re-
 209 trieval algorithm (N2 and D2) at the pixel level. When creating a real SST
 210 product, NEdT varies as a function of both channel and brightness temper-
 211 ature. For N2 retrievals in the example provided, this invariant uncertainty
 212 value is 0.11 K and for D2 retrievals 0.25 K. Uncertainties in gridded aver-
 213 ages reduce by $\frac{1}{\sqrt{n}}$ giving uncertainty estimates of 0.02 K and 0.05 K for N2
 214 and D2 retrievals over fully observed grid cells. In practice, many grid cells
 215 in Level 3 products are not fully observed due to cloud cover. This reduces
 216 the number (n) of observations available and increases the uncertainties from
 217 random effects. This is illustrated in panels 7 and 8 of Figure 1 for N2 and
 218 D2 retrievals. A cloud mask has been superimposed on the simulated data at
 219 the per-pixel level and uncertainties propagated into the 5x5 pixel product.
 220 Observing only part of a given grid cell additionally introduces sampling un-
 221 certainty, discussed briefly in Section 3.4 and more fully in the companion
 222 paper (Bulgin et al., 2016).

223 3.2. Locally Systematic Effects

224 Uncertainties from locally systematic effects arise from ambiguities in or
225 limitations of the SST retrieval algorithm. Coefficient based retrievals for
226 the ATSR instruments in Phase 2 of the SST CCI will use coefficients from
227 the ATSR Reprocessing for Climate (ARC) project. These are calculated
228 based on radiative transfer simulations which cover a comprehensive range of
229 surface and atmospheric conditions (Embury and Merchant, 2012; Embury
230 et al, 2012). Locally systematic effects therefore vary on synoptic scales
231 consistent with changes in atmospheric conditions.

232 We can characterise the uncertainties arising from locally systematic ef-
233 fects in the retrieval scheme using simulation studies. To do this, we take a
234 ‘true’ SST field from Numerical Weather Prediction (NWP) data and simu-
235 late the associated brightness temperatures globally as would be observed by
236 the AATSR instrument using the RTTOV radiative transfer model. We can
237 then use these simulated brightness temperatures as input into our retrieval
238 scheme, comparing our retrieved SST with the ‘true’ SST eg. (Merchant
239 et al., 2009). For any given scene, we can plot the retrieval error field using
240 this methodology as shown in Figure 2. The contour lines in the top pan-
241 els show atmospheric pressure and in the bottom two panels TCWV with
242 the spatial distribution of the error field consistent with synoptic scales of
243 pressure in hPa and total column water vapour (TCWV) in kgm^2 variability.
244 However, features in the SST error field are not simply linked to TCWV
245 distributions, since we see that a single contour line can run through re-
246 gions of both positive and negative errors. The ARC retrieval coefficients are
247 banded by TCWV and the observed errors are not simply a bias that can

248 be removed from the retrieval. Uncertainty arising from these error effects
 249 is characterised in the retrieval as a function of TCWV consistent with the
 250 coefficient banding. Panels in the left and right in Figure 2 show the SST
 251 retrieval error fields for different days, which vary in time as well as space on
 252 synoptic scales.

253 Within the retrieval scheme, uncertainties are calculated as the standard
 254 deviation of the error distributions from the simulated data, taking the dif-
 255 ferences between the ‘true’ and retrieved SSTs. This is the fitting error of
 256 the regression when the coefficients are applied to the simulated data used
 257 to generate the coefficients. Figure 3 shows the uncertainties as a function of
 258 TCWV for retrievals using different channel combinations at different view-
 259 ing geometries. For the N2 retrieval using two channels (11 and 12 μm) the
 260 uncertainties increase as a function of TCWV, flattening at higher TCWV’s
 261 above 45 kg m^{-2} . With the addition of information from multiple viewing
 262 angles (0-22° and 52-55°) locally systematic uncertainties are significantly
 263 reduced to ~ 0.1 K or lower.

264 Figure 3 also shows the uncertainty from uncorrelated effects as a func-
 265 tion of TCWV for different channel combinations. Comparing single-view
 266 retrieval uncertainties with dual-view uncertainties, the dual-view capability
 267 reduces the systematic uncertainty at the expense of the increased retrieval
 268 noise. Uncertainties from uncorrelated effects are dependent on both the
 269 NEdT for a given channel combination and the coefficients. For the N2 and
 270 D2 retrievals large weights are assigned to the 11 and 12 μm channels which
 271 magnifies the uncorrelated uncertainty. ARC coefficients are tuned to assume
 272 NEdTs of 0.01 K (smaller than actual values) as they are designed to produce

273 SST products at 0.1° resolution. This has the effect of reducing locally sys-
274 tematic uncertainties at the cost of increased uncorrelated uncertainties as
275 these decrease as a function of $1/\sqrt{n}$ when calculating the gridded product.

276 Many SST retrievals also use information from the $3.7\ \mu\text{m}$ channel at
277 night. The consequence of adding this third channel to the retrieval (results
278 not shown) reduces uncertainty from locally systematic effects to $\sim 0.1\ \text{K}$ or
279 lower, with larger uncertainties for drier atmospheres. As TCWV increases,
280 the 11 and $12\ \mu\text{m}$ channels become less sensitive to the surface whilst the 3.7
281 μm channel remains relatively transparent. SSTs in regions of high TCWV,
282 close to the equator also show less variability which may improve the fit of
283 the retrieval to the training data. For the uncertainties due to uncorrelated
284 effects, including the $3.7\ \mu\text{m}$ channel in the retrieval results in smaller weights
285 for the 11 and $12\ \mu\text{m}$ channels reducing the noise amplification.

286 *3.3. Large Scale Systematic Effects*

287 Other effects can cause SST errors that are correlated on larger scales.
288 For brevity, the uncertainty associated with unknown errors correlated on
289 large scales is hereafter referred to as “systematic uncertainty”. (It is taken
290 for granted that any ‘known’ or ‘estimated’ systematic errors have been ad-
291 dressed i.e., that any general bias has been quantified and subtracted from
292 data. The systematic uncertainty therefore quantifies the degree of doubt in
293 the measurements associated with what might be termed ‘residual biases’.)

294 All satellite sensors are calibrated prior to launch to a pre-defined stan-
295 dard. The required accuracy for SST measurements from space for climate
296 applications is $0.1\ \text{K}$ (Ohring et al, 2005). In some cases the SST algorithm
297 itself is capable of adjusting for some of the systematic errors in calibra-

tion, for example an SST retrieval algorithm that fits regression coefficients to buoys directly will correct for some of the calibration biases as part of the fitting process. This process will also introduce an additional source of uncertainty from unknown errors in the buoy measurement. The buoy data are point measurements at depth whereas the satellite observations are area measurements of skin temperature. If the sensor is poorly characterised this additional uncertainty term can be smaller than the systematic calibration bias. Thermal channels on some sensors seem in practice to have a BT calibration accuracy of 0.1 K, judging by the SST accuracy achievable using radiative transfer-based coefficients.

The sensor having been calibrated to a certain level, there remain smaller errors, within the specified calibration accuracy, that are unknown. These may vary systematically with scene temperature, general instrument temperature, the thermal state of the on-board calibration target, the temperature of the detectors, the illumination of the sensor on the space-craft by the Sun, and potentially with many other factors. Sometimes, these effects are sufficiently evident in flight that they can be diagnosed and corrected for (Cao et al., 2005; Yu et al., 2012; Wang and Cao, 2008; Mittaz and Harris, 2011; Mittaz et al., 2013). There may be a gradual evolution of such systematic calibration effects over time, as the sensor ages, and/or as the platform orbit drifts, changing the illumination and thermal cycling of the sensor.

Where satellite datasets are reprocessed, there may be some effort to harmonise the BTs across different sensors. “To harmonise” here means to reconcile the calibration of the observed BTs given the known differences between the sensors; it does not mean that the BTs would be the same for

323 two sensors viewing the same scene; it does mean that the differences would
324 be traceable to known instrumental differences, such as different spectral re-
325 sponse functions. The adjustments made to BTs in the light of harmonisation
326 have their own associated uncertainty, and this also is likely to be system-
327 atic as defined here. Overall, harmonisation is intended to reduce systematic
328 effects, particularly relative errors between sensors.

329 It is possible in principle to estimate the systematic uncertainty associated
330 with calibration. There are two possible approaches. The first is to exploit
331 the pre-flight calibration information where an analysis of the potential cal-
332 ibration errors has been made. Where such information is available in suffi-
333 cient detail in the public domain, it can form the basis of an uncertainty bud-
334 get. The second is to exploit near-coincident observations in space between
335 different sensors. Having accounted for instrumental characteristics, differ-
336 ences in matched observations can be used to adjust a less-well-calibrated
337 sensor to a better-calibrated sensor. These adjustments have a quantifiable
338 statistical uncertainty, which then provides an estimate of the magnitude of
339 the post-correction systematic uncertainty eg. (Goldberg, 2007).

340 In general, however, calibration uncertainty is not well quantified and
341 propagation of such information into the systematic uncertainty in SST has
342 not been undertaken, to our knowledge. Arguably, for SSTs generated opera-
343 tionally for use in numerical weather prediction and real-time oceanography,
344 it has not been necessary. However, in the context of developing repro-
345 cessed SST datasets for climate applications, it is an area that needs to be
346 developed. Climate data records require justified uncertainty estimates, par-
347 ticularly estimates of their multi-decadal stability, which implies a detailed

348 engagement with understanding and propagating uncertainty from system-
349 atic effects throughout the record (Minnett and Corlett, 2012). A metrology
350 (science of measurement) of Earth Observation needs to be developed, to
351 bring relevant metrological principles for developing traceable chains of un-
352 certainty to bear in the context of historic satellite missions.

353 *3.4. Sampling Uncertainties*

354 Many users of SST data require gridded products with SST specified as a
355 mean value across the space and time represented by the grid cell. Often grid
356 cells are not fully observed, typically in infrared measurements due to cloud
357 cover, but also in the case of corrupted data or problems with the retrieval
358 process. Data may also be removed from the subsample by conservative cloud
359 detection schemes which can mask clear-sky pixels. The mean SST across
360 the observed pixels may differ from the mean SST across all pixels in the
361 grid cell introducing sampling uncertainty.

362 We cannot explicitly calculate the difference between the SST across the
363 full grid cell and the SST in the available subsample within the retrieval as
364 we do not know the SST of the unsampled pixels. We can however model the
365 sampling uncertainty associated with this process using fully clear-sky data
366 extracts, and we do this as a function of the percentage of the total number
367 of pixels available in the subsample and the standard deviation of the SST
368 in the available pixels.

369 The full details of the derivation of the sampling uncertainty model are
370 provided in the companion paper (Bulgin et al., 2016). Here we provide only
371 a brief overview, for completeness of the discussions in this paper. In Bulgin
372 et al. (2016) we parameterise sampling uncertainty using a cubic function in

the form $(ap^3 + bp^2 + cp + d)$ where a , b , c and d are coefficients dependent on the SST standard deviation in the subsample, and p is the percentage of clear-sky pixels within a given grid cell. This model is therefore applicable to any retrieval scheme with data at the same spatial scale provided that the noise contribution to the SST standard deviation has been subtracted.

3.5. *Other effects contributing to uncertainty*

The propagation of the effects of radiometric noise and the analysis of locally systematic uncertainty discussed has assumed the context of normal clear-sky conditions for each SST retrieval. This neglects the fraction of retrievals that will in practice be made under unusual conditions. These are principally retrievals made for pixels whose classification as clear-sky-over-seawater is doubtful, but which have nonetheless passed the cloud screening process. At present, we have no method for estimating this in the uncertainty budget.

The first case to consider is ‘residual’ unscreened cloud contamination. Clouds escape detection if they are sufficiently small and low (warm) or sufficiently optically thin (e.g., some cirrus). In these cases they can nonetheless affect BTs at the level of several tenths of kelvin. The corresponding impact on SST depends on how different the cloud impacts on BTs are from the impact of increased water vapour in the atmosphere (which the retrieval algorithms are adapted to deal with). The probability of such cases is considered to be greater around the edges of areas correctly identified as cloudy. Note that both the distribution of BT modification by cloud-contamination in pixels falsely considered to be clear sky, as well as the frequency of failure to detect are dependent on the cloud screening system. One could envisage

398 that simulation of a representative range of cloudy situations be carried out
399 to generate such information, but to our knowledge, this has not been done.
400 Given these pieces of information, assessment of the contribution to SST un-
401 certainty could be undertaken by error propagation methods similar to those
402 described earlier. At present, however, the contribution of this effect to SST
403 uncertainty is not estimated.

404 The second case to consider is atmospheric aerosol of a form and optical
405 depth outside the range of circumstances for which the retrieval coefficients
406 are designed. Again, to the degree that the aerosol affects the BTs differently
407 to water vapour [e.g., Merchant et al. (2006)], SST errors may be induced
408 of unknown size. While aerosol events trigger cloud detection if the optical
409 depths are sufficiently great, there is a regime where SST retrievals can be
410 affected, the effect in most cases being to make the retrieved SST too cold.
411 Again, the contribution of this effect to SST uncertainty is not estimated.

412 The third case relates to sea ice being present within the pixel for which
413 SST is retrieved. If the ice is not too cold and is relatively dark (circumstances
414 that often go together in the formation of new ice), the ice may not be
415 detected. Similar considerations apply as to missed residual cloud or aerosol,
416 and this contribution to uncertainty again is not presently estimated.

417 There are a number of further effects contributing to SST uncertainty that
418 are neglected in the SST CCI uncertainty model. These include differences
419 in the instantaneous field of view for channels of different wavelength, and
420 local to regional variations in trace gas concentrations.

421 4. Validation of the Uncertainty Budget

422 Having constructed an initial uncertainty budget for remotely sensed
423 SSTs independently of in-situ data, we can now use these in-situ data to
424 validate our uncertainties (as well as the retrieved SST). In Section 3, we
425 characterised two quantifiable components of uncertainty relating to SSTs
426 calculated from satellite data at a pixel level (a random component due to
427 noise in the data and a locally systematic component arising from uncertain-
428 ties varying on a synoptic scale within the retrieval) from which we construct
429 our initial uncertainty budget. We validate this budget using data from the
430 AATSR instrument spanning four years (2006 - 2009 inclusive) considering
431 both the N2 and D2 retrievals. The data used in the validation are taken
432 from the SST CCI multi-sensor match-up system (MMS) (Corlett et al.,
433 2014) where drifting buoy and satellite observations are matched globally
434 under clear-sky conditions (Corlett et al., 2014).

435 Matches are filtered to include only the closest in-situ match in time to
436 the satellite observation and to check the quality of the in-situ data. Matches
437 can have a maximum time difference of 4 hours and maximum spatial separation
438 of 10 km. Bad quality in-situ data are removed based on the following
439 criteria 1) absolute difference between NWP and in-situ SST greater than
440 10 K, 2) standard deviation of the in-situ SST history greater than 5 K and
441 3) standard deviation of the in-situ latitude history greater than 10 degrees.
442 Validation of satellite data using in-situ data necessitates a comparison be-
443 tween a point measurement and the satellite footprint. There are uncertain-
444 ties in this process arising from comparing two different types of observation
445 and geolocation errors in both the satellite and in-situ data. The filtering is

446 therefore necessary to minimise both spatial and temporal separation of the
447 satellite and in-situ observations (Minnett, 1986; Donlon et al., 2002; Corlett
448 et al., 2006).

449 For each match up, the uncertainties in the retrieved SST are calculated
450 as follows. The noise in a given observation is a function of both the channels
451 and associated brightness temperature, and is calculated by monitoring in-
452 orbit blackbody temperature signals (Smith et al., 2012). For AATSR, the
453 NEdT is fairly consistent throughout the lifetime of the mission. These NEdT
454 values are used to calculate the uncertainty due to uncorrelated effects
455 at L2 using the methodology presented in Section 3. The uncertainty from
456 locally systematic effects is quantified as a function of the TCWV consistent
457 with the banding of the retrieval coefficients. In both cases the uncertainties
458 are then propagated into the gridded product for validation of data in L3
459 format. For the gridded products, a sampling uncertainty is also calculated
460 due to the presence of cloud preventing observation of all pixels within a
461 given grid cell (Bulgin et al., 2016). This is an additional uncertainty due
462 to uncorrelated effects that is introduced in the gridding process. At both
463 the per pixel and gridded scales the uncertainty components are added in
464 quadrature to give a total uncertainty.

465 The validation data for the N2 and D2 pixel level retrievals are shown
466 in the top two panels of Figure 4. Here we plot the standard deviation of
467 the SST difference (retrieval minus drifting buoy) against the SST retrieval
468 uncertainty which we have calculated independently represented by the thin
469 black lines in Figure 4. The dashed lines indicate the uncertainty model we
470 would expect to see based on retrieved SST minus drifting buoy differences.

471 There is a lower limit on this model of ± 0.15 K which represents the
 472 uncertainty in the drifting buoy measurements. We chose the time period
 473 of 2006-2009 inclusive for our validation as the drifting buoy uncertainty
 474 has been stable at around 0.15 K over this period (Lean and Saunders ,
 475 2012). The blue line on the plots indicate the median difference between the
 476 retrieved and in-situ SST across all match-ups in each uncertainty bin (width
 477 0.02 K). The standard error in this value is represented by the error bars.
 478 Red lines at the end of the black bars indicate the statistical uncertainty in
 479 the calculated standard deviation and are visible primarily for bins where
 480 the number of contributing cases is small.

481 For the N2 pixel level data we find that our uncertainty estimates closely
 482 match the expected uncertainty model below a total uncertainty of 0.25 K.
 483 Above this threshold, our estimated retrieval uncertainties are too high: a
 484 better fit would be obtained if the bins shifted to lower estimated uncertainty
 485 values. For the D2 retrieval, we see that our uncertainties calculated within
 486 the retrieval process show excellent agreement with the expected uncertainty
 487 model. At a per-pixel level the dominant terms in the uncertainty budget
 488 come from the uncorrelated and locally systematic effects, assuming that a
 489 good cloud detection algorithm is used. Therefore the validation indicates
 490 that our estimate of these components is well constrained.

491 We also consider the validation of uncertainties for gridded N2 and D2
 492 retrievals across a 5x5 pixel domain approximately corresponding to 0.05° .
 493 In this case we also include the sampling uncertainty component in our initial
 494 uncertainty budget (Bulgin et al., 2016). The results for this validation are
 495 shown in the bottom two panels of Figure 4. When considering gridded

496 data we find a larger range of estimated uncertainty than for the per pixel
497 data. This is because SST varies across the gridded domain, and for cells
498 that are not well sampled, the uncertainty on the mean SST increases. For
499 the N2 gridded data we see a similar pattern to the N2 per pixel data with
500 uncertainties being slightly overestimated. For the D2 gridded retrieval the
501 overall uncertainties are smaller, but we underestimate the total uncertainty.

502 5. Discussion

503 Overall, we see that our independent uncertainty estimates show good
504 agreement with validation data using in-situ drifting buoy measurements.
505 The best agreement is for the D2 retrieval at a per-pixel level. For the N2
506 retrievals we see a similar over-estimation of uncertainties above 0.2-0.25
507 K in both the pixel level and gridded products. The uncertainty budget
508 constructed is based on the errors that we currently have the capability to
509 estimate and propagate through the retrieval. Some of the sources of error
510 discussed in the earlier sections such as residual unscreened cloud contam-
511 ination, failure to detect clear-sky pixels and aerosol are not yet included.
512 These may be larger across a gridded domain if they affect multiple pixels.

513 In this validation, the estimation of large scale systematic uncertainties
514 has also been excluded, but in the SST CCI Version 1 products this is set
515 to an invariant value of 0.1 K per pixel as a best estimate of the magnitude
516 of this component, and then added in quadrature to the uncertainty budget
517 (Merchant et al., 2014).

518 Although at present the uncertainty budget can not be fully constrained
519 due to the limitations described in the Section 3, we are able to characterise

520 well the components resulting from random, locally systematic and sampling
 521 effects across a range of retrievals for the ATSR instruments as evidenced
 522 by the good validation statistics. On the relatively short spatial and tem-
 523 poral scales (pixel to gridded averages at 0.1° and instantaneous measure-
 524 ments) the uncertainties from uncorrelated and local systematic effects are
 525 the dominant terms in the uncertainty budget. The contributions from the
 526 ‘missing’ components are therefore relatively small under these SST retrieval
 527 conditions. Empirical systematic effects (biases) are within the estimated un-
 528 certainties and these uncertainties can successfully distinguish more and less
 529 certain SSTs. The approach outlined in this paper has a wider application to
 530 coefficient based SST retrievals using other algorithms and data from other
 531 instruments. If the data provider or user knows the NEdT distribution for
 532 each channel used in the retrieval they can propagate this through the algo-
 533 rithm to obtain the uncertainty due to uncorrelated effects in the retrieved
 534 SST. Data providers can use simulation studies to characterise the locally
 535 systematic uncertainty in their retrieval scheme, and the sampling model is
 536 applicable to any SST retrieval on the same spatial scales as discussed in
 537 this paper provided that the uncertainty due to noise is removed first. Provi-
 538 sion of uncertainty information as part of the retrieval process then enables
 539 validation of these uncertainty estimates, as well as the SST, using in-situ
 540 data.

541 Figure 5 maps mean uncertainty estimates for 2010. The uncertainty
 542 maps show the square root of the mean of the error variance across all days
 543 with observations. Where more than one observation is available for a given
 544 day, the smallest error variance has been used. The uncertainty from uncorre-

lated effects (a) contains the noise and sampling uncertainty components and when added to the uncertainty due to locally systematic effects (b) in quadrature, produces the total uncertainty map (c). Total uncertainties typically range between 0.1-0.25 K globally, with the highest values predominantly in equatorial regions and some northern hemisphere high latitudes. The uncertainty due to uncorrelated effects is the larger contributor to this signal, and in these regions scattered or patchy cloud cover increases sampling uncertainties. Figure 5 (d) also shows the ratio of the retrieved SST variability to the uncertainty, calculated by dividing the standard deviation of the SST in an given location over the whole of 2010 by the total uncertainty. The highest ratios are seen in mid-latitude regions where SSTs show greater seasonal variation.

6. Conclusions

In this paper we present a framework for the provision of uncertainty estimates in coefficient based SST retrieval from satellite data, based on propagation of noise, simulation of noise-free retrieval errors, and empirical characterisation of sampling effects. The uncertainty estimates can be validated in their own right, in addition to validating the retrieved SST. We provide a detailed discussion of different sources of uncertainty in the SST retrieval and how to propagate these through the retrieval process. We derive three uncertainty components here and in the companion paper; uncertainties due to uncorrelated, locally systematic and sampling effects. We apply our derivation to AATSR data within the context of the SST CCI project and find that our uncertainties validate well against in-situ data for both per

569 pixel and gridded products, and for two different retrieval algorithms.

570 **7. Acknowledgements**

571 The work undertaken in this paper was funded by the European Space
572 Agency Sea Surface Temperature Climate Change Initiative project.

573 **References**

- 574 Anding, D. & Kauth, R. (1970). Estimation of sea surface temperature from
575 space. *Remote Sensing of Environment*, 1, 217-220.
- 576 Barton, I. J. (1998). Improved techniques for the derivation of sea surface
577 temperatures from ATSR data. *Journal of Geophysical Research*, 103,
578 8139-8152.
- 579 Brohan, P. & Kennedy, J. J., & Harris, I., & Tett, S. F. B., & Jones, P. D.
580 (2006). Uncertainty estimates in regional and global observed temperature
581 changes: A new data set from 1850. *Journal of Geophysical Research*, 111.
- 582 Bulgin, C. E., & Embury, O., & Merchant, C. J. (2016). Sampling Uncer-
583 tainty in Global Area Coverage (GAC) and Gridded Sea Surface Temper-
584 ature Products. *Remote Sensing of Environment*.
- 585 Coa, C., & Sullivan, J. & Maturi, E. & Sapper, J. (2004). The effect of orbit
586 drift on the calibration of the 3.7 μm channel of the AVHRR onboard
587 NOAA-14 and its impact on night-time sea surface temperature retrievals.
588 *International Journal of Remote Sensing*, 25, 975-986.

589 Casey, K. S., & Cornillon, P. (1998). A Comparison of Satellite and In Situ-
590 Based Sea Surface Temperature Climatologies. *Journal of Climate*, 12,
591 1848-1863.

592 Castro, S., & Cornillon, P., & Gentemann, C., & Jessup, A., & Hacker, P., &
593 Kaplan, A., & Lindstrom, E., & Maturi, E., & Minnett, P., & Reynolds,
594 R. (2010). Sea Surface Temperature Error Budget: White Paper. Interim
595 Sea Surface Temperature Science Team (ISSTST), pages 44.

596 Corlett, G. K., & Barton, I. J., & Donlon, C. J., & Edwards, M. C., & Good,
597 S. A., & Horrocks, L. A., & Llewellyn-Jones, D. T., & Merchant, C. J., &
598 Minnett, P. J., & Nightingale, T. J., & Noyes, E. J., & O'Carroll, A. G.,
599 & Remedios, J. J., & Robinson, I. S., & Saunders, R. W., & Watts, J. G.
600 (2006). The accuracy of SST retrievals from AATSR: An initial assessment
601 through geophysical validation against in situ radiometers, buoys and other
602 SST data sets.. *Advances in Space Research*, 37, 764-769.

603 Corlett, G., & Atkinson, C., & Rayner, N., & Good, S., & Fiedler, E., &
604 McLaren, A., & Hoeyer, J., & Bulgin, C. E. (2014). Product Validation
605 and Intercomparison Report (PVIR) SST_CCI-PVIR-UoL-001, pp. 14.

606 Corlett, G. K., & Merchant, C. J. & Minnett, P. J. & Donlon, C. J.
607 (2014). Assessment of Long-Term Satellite Derived Sea Surface Tem-
608 perature Records. In: Zibordi, G., Donlon, C. J. and Parr, A. C. (eds.)
609 Optical radiometry for ocean climate measurements. Experimental meth-
610 ods in the physical sciences 47(47).. Academic Press, pp. 639-674. ISBN
611 9780124170117 doi: 10.1016/B978-0-12-417011-7.00015-5

- 612 Deschamps, P. Y., & Phulpin, T. (1980). Atmospheric correction of infrared
613 measurements of sea surface temperature using channels at 3.7, 11 and 12
614 μm . *Boundary Layer Meteorology*, 18, 131-143.
- 615 Donlon, C. J., & Minnett, P. J., & Gentemann, C., & Nightingale, T. J.,
616 & Barton, I. J., & Ward, B., & Murray, M. J. (2002). Toward Improved
617 Validation of Satellite Sea Surface Skin Temperature Measurements for
618 Climate Research. *Journal of Climate*, 15, 353-369.
- 619 Donlon, C., & Robinson, I., & Casey, K. S., & Vazquez-Cuevo, J., & Arm-
620 strong, E., & Arino, O., & Gentemann, C., & May, D., & LeBorgne, P.,
621 & Piollè, J., & Barton, I., & Beggs, H., & Poulter, D. J. S., & Merchant,
622 C. J., & Bingham, A., & Heinz, S., & Harris, A., & Wick, G., & Emery,
623 B., & Minnett, P., & Evans, R., & Llewellyn-Jones, D., & Mutlow, C., &
624 Reynolds, R. W., & Kawamura, H., & Rayner, N. (2007). The Global
625 Ocean Data Assimilation Experiment High-resolution Sea Surface Tem-
626 perature Pilot Project. *Bulletin of the American Meteorological Society*,
627 1197-1213.
- 628 Embury, O., & Merchant, C. J., & Corlett, G. K (2012). A reprocessing for
629 climate of sea surface temperature from the along-track scanning radiome-
630 ters: initial validation, accounting for skin and diurnal variability. *Remote*
631 *Sensing of Environment*, 116, 62-78.
- 632 Embury, O., & Merchant, C. J. (2012). A reprocessing for climate of sea
633 surface temperature from the along-track scanning radiometers: A new
634 retrieval scheme. *Remote Sensing of Environment*, 116, 47-61.

Embury, O., & Merchant, C. J. & Fillipiak, M. J. (2012). A reprocessing
for climate of sea surface temperature from the along-track scanning ra-
diometers: basis in radiative transfer. *Remote Sensing of Environment*,
116, 32-46.

Folland, C. K., & Rayner, N. A., & Brown, S. J., & Smith, T. M., & Shen,
S. S. P., & Parker, D. E., & Macadam, I., & Jones, P. D., & Jones, R. N.,
& Nicholls, N., & Sexton, D. M. H. (2001). Global temperature change
and its uncertainties since 1861. *Geophysical Research Letters*, 28, 13,
2621-2624.

GHR SST Science Team (2010). The Recommended GHR SST Data Speci-
fication (GDS) Revision 2.0 Technical Specifications. Available from the
GHR SST International Project Office, <http://www.ghrsst.org>, pp. 120

Goldberg, M (2007) Global space-based inter-calibration system (GSICS)
Proc. SPIE 6684, Atmospheric and Environmental Remote Sensing
Data Processing and Utilization III: Readiness for GEOSS, 668402,
<http://dx.doi.org/10.1117/12.735246>.

Hollman, R., & Merchant, C. J., & Saunders, R., & Downy, C., & Buchwitz,
M., & Cazenave, A., & Chuvieco, E., & Defourny, P., & de Leeuw, G.,
& Forsberg, R., & Holzer-Popp, T. & Paul, F. & Sandven, S. & Sathyen-
dranath, S. & van Roozendaal, M. & Wagner, W. (2013). The ESA Climate
Change Initiative: Satellite Data Records for Essential Climate Variables.
Bulletin of the American Meteorological Society, 94, 1541-1552.

Joint Committee for Guides in Metrology (2008). Evaluation of mea-

658 surement data - Guide to the expression of uncertainty in measure-
659 ment. Bureau International des Poids et Mesures, available from
660 www.bipm.org/en/publications/guides/, pp. 130.

661 Jones, P. D., & Osborn, T. J., & Briffa, K. R. (1997). Estimating Sampling
662 Errors in Large-Scale Temperature Averages. *Journal of Climate*, *10*,
663 2548–2568.

664 Kennedy, J. J. (2014). A review of uncertainty in in situ measurements and
665 data sets of sea surface temperature. *Reviews of Geophysics*, *52*, 1-32.

666 Kilpatrick, K. A., & Podestá, G. P., & Evans, R. (2001). Overview of the
667 NOAA/NASA advanced very high resolution radiometer Pathfinder al-
668 gorithm for sea surface temperature and associated matchup database.
669 *Journal of Geophysical Research*, *106*, 9179-9197.

670 Lean, K., & Saunders, R. W. (2012). Validation of the ATSR Reprocessing
671 for Climate (ARC) Dataset Using Data from Drifting Buoys and a Three-
672 Way Error Analysis. *Journal of Climate*, *26*, 1, 4758-4772.

673 Le Borgne, P., & Roquet, H., & Merchant, C. J. (2011). Estimation of sea
674 surface temperature from the Spinning Enhanced Visible and Infra Red
675 Imager, improved using numerical weather prediction. *Remote Sensing of*
676 *Environment*, *115*, 1, 55-65.

677 May, D. A., & Parmeter, M. M., & Olszewski, D. S., & McKenzie, B. D.
678 (1997). Operational Processing of Satellite Sea Surface Temperature Re-
679 trievals at the Naval Oceanographic Office. *Bulletin of the American Me-*
680 *teorological Society*, 397-407.

- 681 Merchant, C. J. (2013). Thermal remote sensing of sea surface temper-
682 ature In: Kuenzer, C. and Dech, S. (eds.) Thermal Infrared Remote
683 Sensing: Sensors, Methods, Applications. Remote Sensing and Digital
684 Image Processing, 17. Springer Netherlands, Dordrecht, pp.287-313. ISBN
685 9789400766389.
- 686 Merchant, C. J., & Embury, O., & Rayner, N. A., & Berry, D. I., & Corlett,
687 G. K., & Lean, K., & Veal, K. L., & Kent, E. C., & Llewellyn-Jones,
688 D. T., & Remedios, J. J., & Saunders, R. (2012). A 20 year independent
689 record of sea surface temperature for climate from Along-Track Scanning
690 Radiometers. *Journal of Geophysical Reserach*, 117, C12013, pp.18.
- 691 Merchant, C. J., & Embury, O., & Le Borgne, P., & Bellec, B. (2006). Saha-
692 ran dust in night-time thermal imagery: detection and reduction of related
693 biases in retrieved sea surface temperature. *Remote Sensing of Environ-*
694 *ment*, 104, 15-30.
- 695 Merchant, C. J., & LeBorgne, P. (2004). Retrieval of sea surface temperature
696 from space based on modeling of infrared radiative transfer: capabilities
697 and limitations. *Journal of Atmospheric and Oceanic Technology*, 21, 11,
698 1734-1746.
- 699 Merchant, C. J., & Harris, A. R., & Roquet, H., & Le Borgne, P. (2009).
700 Retrieval characteristics of non-linear sea surface temperature from the Ad-
701 vanced Very High Resolution Radiometer. *Geophysical Research Letters*,
702 36, 17, 1-5 .
- 703 Merchant, C. J., & Embury O. & Roberts-Jones, J. & Fiedler, E. & Bulgin,

704 C. E. & Corlett, G. K. & Good, S. & McLaren, A. & Rayner, N. & Donlon,
705 C. (2014). Sea surface temperature datasets for climate applications from
706 Phase 1 of the European Space Agency Climate Change Initiative (SST
707 CCI). *Geoscience Data Journal*, 1, 2, 179-191.

708 McMillan, L. M., & Crosby, D. S. (1984). Theory and Validation of the Mul-
709 tiple Window Sea Surface Temperature Technique. *Journal of Geophysical*
710 *Research*, 89, C3, 3655-3661.

711 Minnett, P. J. (1991). Consequences of sea surface temperature variability
712 on the validation and applications of satellite measurements. *Journal of*
713 *Geophysical Research*, 96, C10, 18475-18489.

714 Minnett, P. J. (1986). A numerical study of the effects of anomalous North
715 Atlantic atmospheric conditions on the infrared measurement of sea surface
716 temperature from space. *Journal of Geophysical Research*, 91, C7, 2156-
717 2202.

718 Minnett, P. J., & Corlett, G. K. (2012). A pathway to generating Climate
719 Data Records of sea-surface temperature from satellite measurements.
720 *Deep Sea Research Part II: Tropical Studies in Oceanography*, 77-80, 44-
721 51.

722 Mittaz, J., & Harris, A. (2011). A Physical Method for the Calibration of the
723 AVHRR/3 Thermal IR Channels. Part II: An In-Orbit Comparison of the
724 AVHRR Longwave Thermal IR Channels on board MetOp-A with IASI.
725 *Journal of Atmospheric Oceanic Technology*, 28, 1072-1087.

- 726 Mittaz, J., & Bali, M. & Harris, A. (2013). The Calibration of Broad
 727 Band IR Sensors: Time variable biases and other issues. *EUMET-*
 728 *SAT Meteorological Satellite Conference, Vienna, Available online at*
 729 *[http://www.eumetsat.int/website/wcm/idc/idcplg?IdcService=GET_FILE](http://www.eumetsat.int/website/wcm/idc/idcplg?IdcService=GET_FILE&dDocName=PDF_CONF_P_S8_11_MITTAZ_V&RevisionSelectionMethod=L)*
 730 *&dDocName=PDF_CONF_P_S8_11_MITTAZ_V&RevisionSelectionMethod=L,*
 731 *pp. 8.*
- 732 Morrissey, M. L., & Greene, J. S. (2009). A theoretical framework for the
 733 sampling error variance for three-dimensional climate averages of ICOADS
 734 monthly ship data. *Theoretical Applied Climatology*, 96, 235-248.
- 735 Ohring, G., & Wielicki, B., & Spencer, R., & Emery, B., & Datla, R. (2005).
 736 Satellite Instrument Calibration for Measuring Global Climate Change.
 737 *Bulletin of the American Meteorological Society*, 1303-1313.
- 738 Rayner, N. A., & Brohan, P., & Parker, D. E., & Folland, C. K., & Kennedy,
 739 J. J., & Vanicek, M., & Ansell, T. J., & Tett, S. F. B. (2006). Improved
 740 Analyses of Changes and Uncertainties in Sea Surface Temperature Mea-
 741 sured In Situ since the Mid-Nineteenth Century: The HadSST2 Dataset.
 742 *Journal of Climate*, 19, 446-469.
- 743 Reynolds, R. W., & Rayner, N. A., & Smith, T. M., & Stokes, D. C., &
 744 Wang, W (2002). An Improved In Situ and Satellite SST Analysis for
 745 Climate. *Journal of Climate*, 15, 1609-1625.
- 746 She, J., & Hoyer, J. L., & Larsen, J. (2007). Assessment of sea surface tem-
 747 perature observational networks in the Baltic Sea and North Sea. *Journal*
 748 *of Marine Systems*, 65, 314-335.

- 749 Smith, D., & Mutlow, C.,& Delderfield, J. , & Watkins, B.,& Mason, G.
750 (2012). ATSR infrared radiometric calibration and in-orbit performance.
751 *Remote Sensing of Environment*, 116, 4-16.
- 752 Wang, L., & Cao, C. (2008). On-Orbit Calibration Assessment of AVHRR
753 Longwave Channels on MetOp-A Using IASI. *IEEE Transactions on Geo-*
754 *sciences and Remote Sensing*, 46, 4005-4013.
- 755 Yu, F., & Wu, X., & Rama Varma Raja, M. K., & Wang, L., & Goldberg,
756 M. (2012). Diurnal and Scan Angle Variations in the Calibration of GOES
757 Imager Infrared Channels. *IEEE Transactions on Geosciences and Remote*
758 *Sensing*, 51, 671-683.

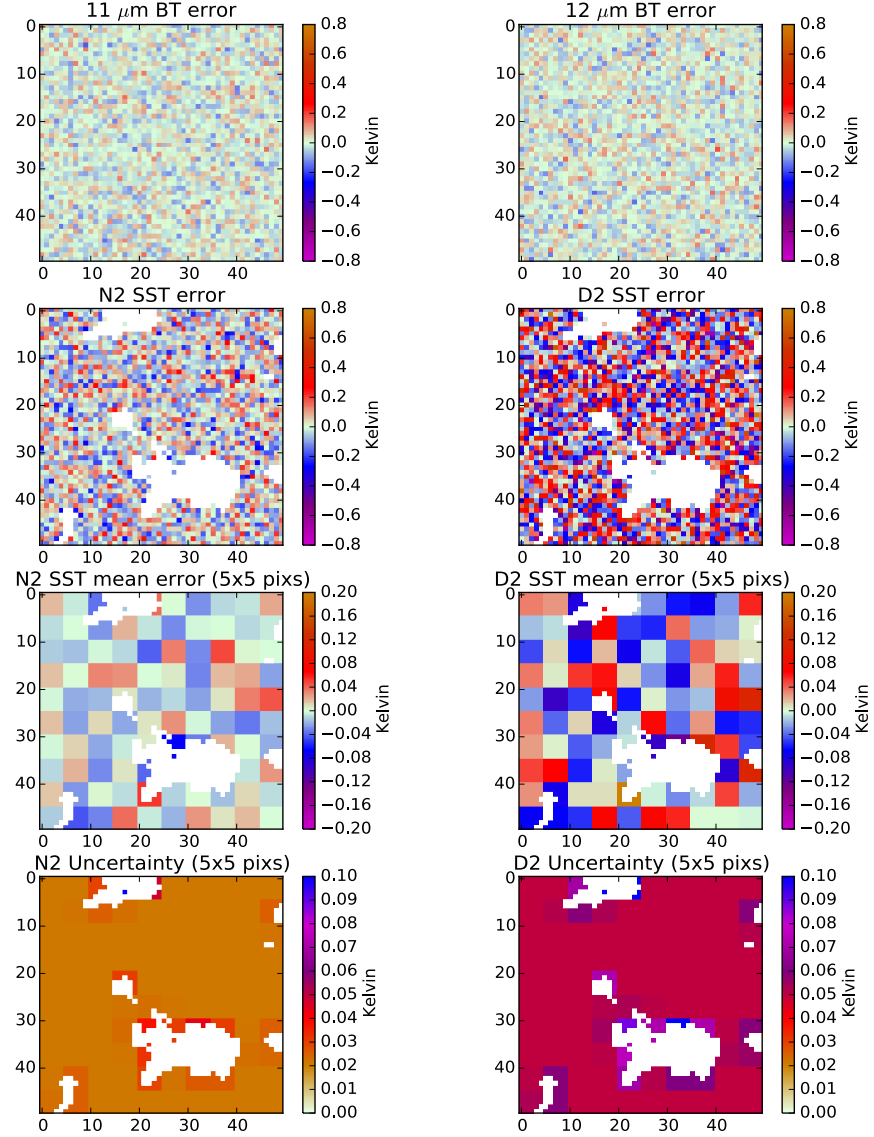


Figure 1: Uncorrelated random errors and uncertainties in brightness temperature observations and SST retrieval. Panels a) and b) show simulated errors in the 11 and 12 μm channels. Panels c) and d) show these errors propagated into SST retrievals for N2 and D2 retrievals. Panels e) and f) show the mean error at a 5x5 pixel resolution with a cloud mask superimposed on the data. Panels g) and h) show the associated uncertainty fields at a 5x5 pixel resolution.

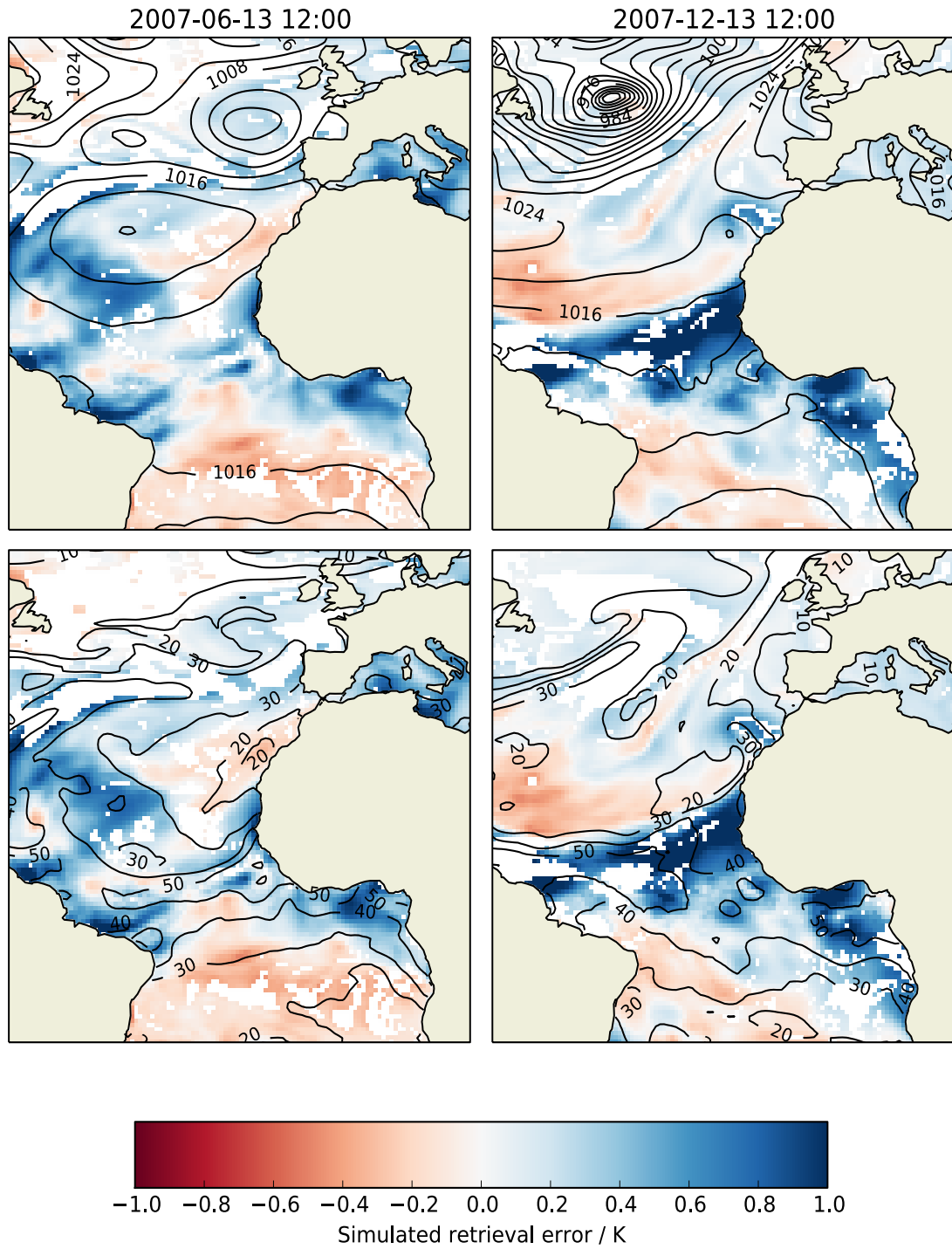


Figure 2: AATSR retrieval errors for two different days from simulation studies (left and right). Plots show the difference between the ‘true’ and retrieved SST field. Plots in the upper panels show pressure contours hPa, and plots in the lower panels TCWV contours kg m⁻².

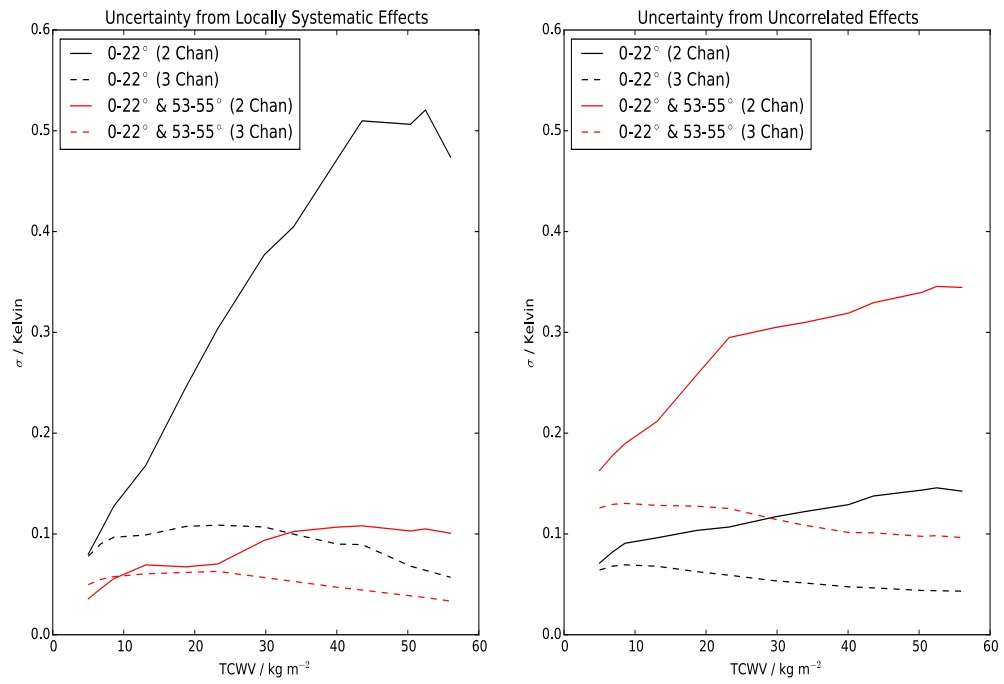


Figure 3: Uncertainties from a) locally systematic and b) uncorrelated effects as a function of total column water vapour for different channel combinations.

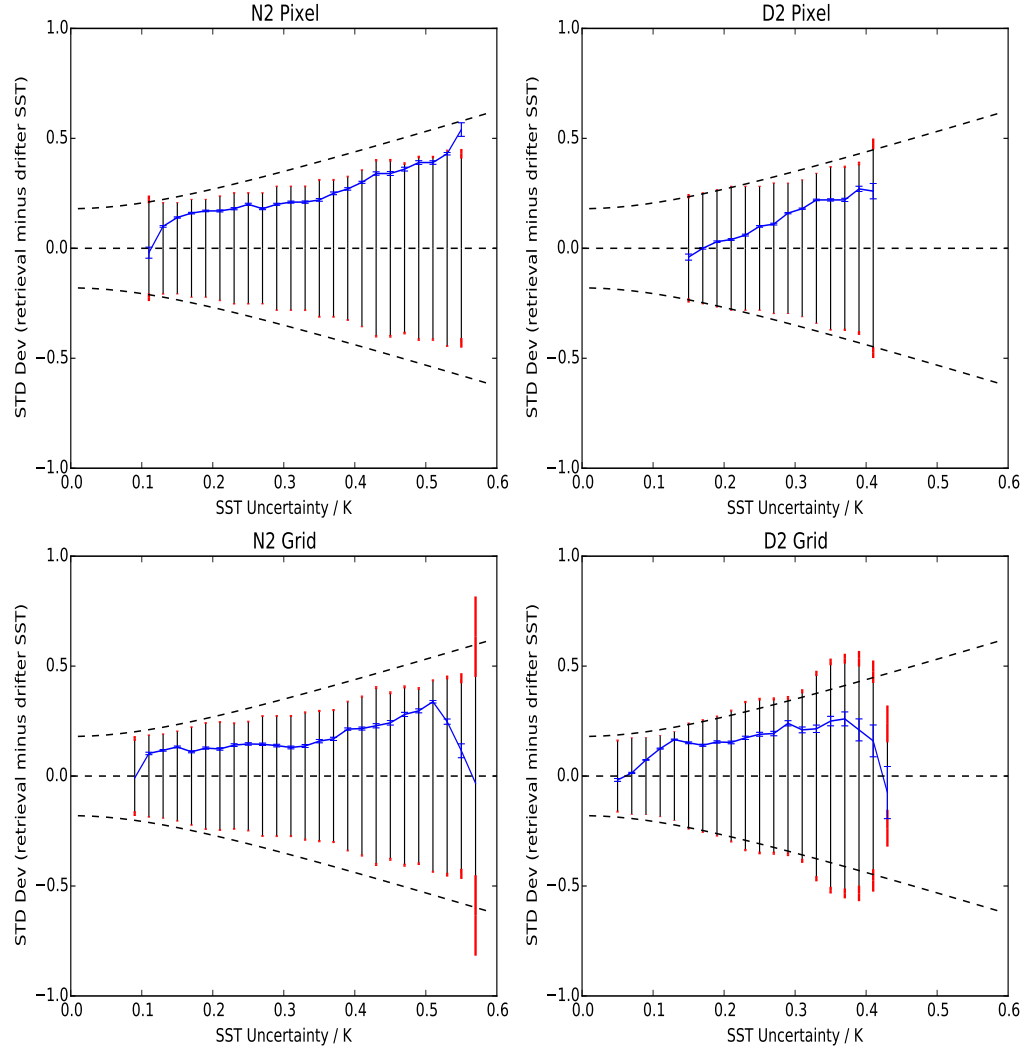


Figure 4: SST uncertainty validation against drifting buoy in-situ data. Top panels show pixel level uncertainties for N2 and D2 retrievals. Bottom panels show grid cell uncertainties (5x5 pixels approximately corresponding to a resolution of 0.05°) for N2 and D2 retrievals. Dashed lines show ideal uncertainty model accounting for uncertainties in the buoy data and geophysical uncertainties arising from a skin to depth comparison and collocation. Solid black lines show one standard deviation of the retrieved minus buoy SST differences, and blue lines the median satellite minus buoy SST difference. Error bars show the standard error in these differences. Uncertainties in the retrieval uncertainty are indicated by red bars at the base and top of the solid black lines.

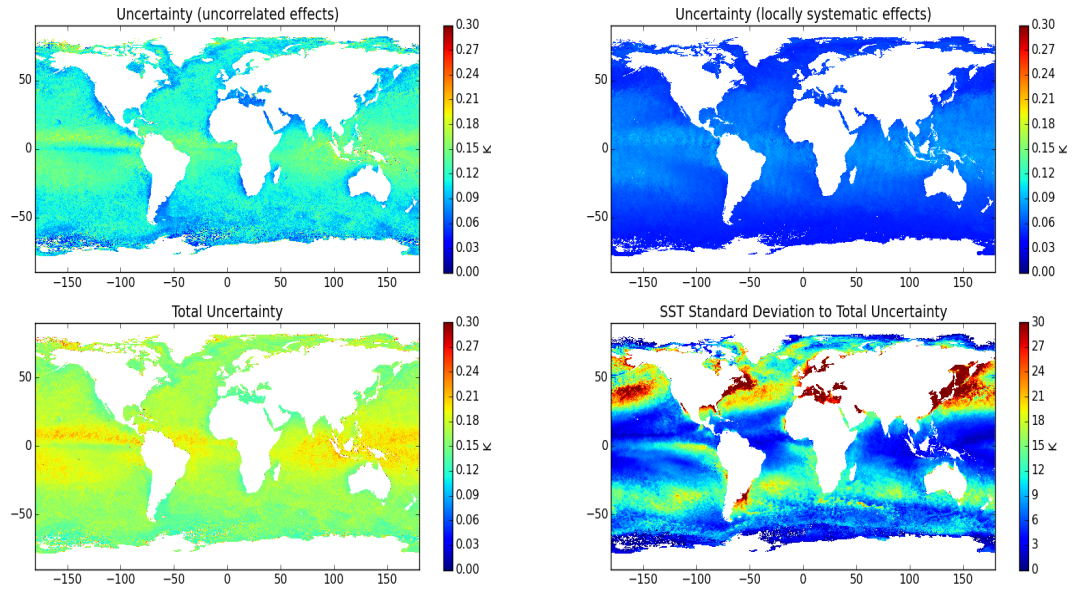


Figure 5: Annual means in SST retrieval uncertainties calculated from AATSR L3C data in 2010. Mean uncertainties are derived by adding all uncertainty observations in a given cell in quadrature, dividing by the number of observations and taking the square root. a) Shows uncertainty due to uncorrelated effects (noise and sampling uncertainty), b) shows noise due to locally systematic effects and c) total uncertainty. d) Shows the ratio of the SST standard deviation over 2010 to the total uncertainty.